TWODML: A Markup Language for Unifying Two-Dimensional Polyacrylamide Gel Electrophoresis Data

*Ravichandran, V.[1], Lubell, J.[*2], Vasquez, G.B.[1], Lemkin, P.[3], Sriram, R.D.[2], Gilliland, G.L.[1]*
[1]Biotechnology Division, [2]Manufacturing Systems Integration Division, National Institute of Standards and Technology, Gaithersburg, MD, USA; [3]Laboratory of Computational and Experimental Biology, National Cancer Institute at Frederick, Frederick, MD, USA

Two-dimensional polyacrylamide gel electrophoresis (2-D PAGE) is a common method used to fractionate and identify proteins. Many technical advancements including automation towards this method lead to accumulation of data. As 2-D PAGE plays a major role in proteomics research, it is necessary to compare data from numerous gels, and many of theses samples are unique and cannot be reproduced. Currently, no data standards exist for 2-D PAGE data to systematically establish the correlation between data from different experiments to allow meaningful comparisons. This raises the difficulty in data exchange and interoperability between different data resources, even though they may have some data in common. Despite this, a wealth of data currently exists and is readily available and applicable to proteomics. Thus, data interoperability requires the development of standard data definitions and transfer protocols. Such standards for databases and data reporting can be applied to 2-D PAGE technology, a critical method employed in proteomics efforts. Comprehensive, structured information about the 2-D PAGE data will aid the deeper understanding of a particular protein. The adoption of common standards and ontologies for the management and sharing of 2-D data is essential and will provide immediate benefit to the proteomics community. Hence, a role for the National Institute of Standards and Technology (NIST) and other similar standards and measurement organizations is emerging in facilitating this development.

Here an approach toward standardizing 2-D PAGE data is proposed in support of developing a globally relevant proteomics consensus in order to provide more efficient database querying and data comparisons through the establishment of the necessary definitions and interdisciplinary reference fields for both the 2-D PAGE community, particularly in the proteomics area, and the clinical and experimental biological research communities, in general. This approach addresses the need for unifying the 2-D PAGE data through a common data repository as well as the need for data standards for interoperability. We propose a common language for sharing electrophoresis experimental data, Two-dimensional Electrophoresis Markup Language (TWODML) that is based on the Extensible Markup Language (XML).